DOCUMENT RESUME

ED 317 026 FL 017 771

AUTHOR Mepham, Michael; Berube, Richard
TITLE Le Logiciel Lemmatiseur: Guide d'utilisation (The Lemmatiseur Software Program: User Manual). Publication K-1.
INSTITUTION Laval Univ., Quebec (Quebec). International Center for Research on Bilingualism.
REPORT NO ISBN-2-89219-180-7
PUB DATE 87
NOTE 49p.
AVAILABLE FROM International Center for Research on Bilingualism, Par Casault-Universite Laval, Quebec GIK 7P4, Canada.
PUB TYPE Reports - Descriptive (141)
LANGUAGE French

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Computational Linguistics; *Computer Software; *Discourse Analysis; Foreign Countries; *Semantics

ABSTRACT

A guide to the use of a computer program for discourse analysis is presented. The program, which derives root-words from other word forms, performs an automated, interactive analysis of words in their immediate context, using a previously developed dictionary program. The program is hierarchically menu-driven. The guide outlines the program's approach in general terms, and then systematically describes its use and application. The latter description includes the program's available options for the preliminary treatment of the corpus, generation and maintenance of the factor databank for analysis, analysis of the corpus, and presentation of results. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED317026

PUBLICATION
K-1

(CIRB)
Centre international de recherche
sur le bilinguisme

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

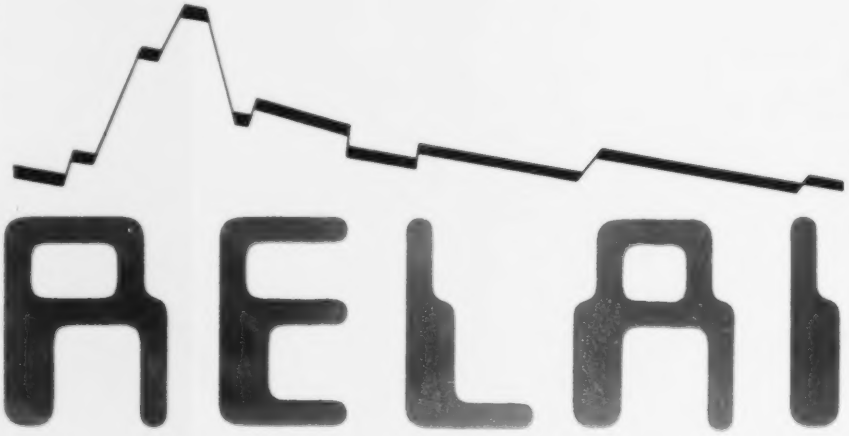
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

L. Laforge

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."



Recherche
en
linguistique
appliquée
à
l'informatique

LE LOGICIEL LEMMATISEUR

Guide d'utilisation

FL017771

Michael Mepham
&
Richard Bérubé

cirb
icrb

2

1987

Michael MEPHAM & Richard BÉRUBÉ

LE LOGICIEL LEMMATISEUR:

Guide d'utilisation

Publication K-1

1987
Centre international de recherche sur le bilinguisme
International Center for Research on Bilingualism
Québec

Le Centre international de recherche sur le bilinguisme est un organisme de recherche universitaire qui reçoit une contribution du Secrétariat d'État du Canada pour son programme de publication.

The International Center for Research on Bilingualism is a university research institution which receives a contribution from the Secretary of State of Canada for its publication programme.

© **1987 CENTRE INTERNATIONAL DE RECHERCHE SUR LE BILINGUISME**
Tous droits réservés. Imprimé au Canada
Dépôt légal (Québec) 4e trimestre 1987
ISBN 2-89219-180-7

PRÉSENTATION DE LA COLLECTION "K"

En inaugurant aujourd'hui, après vingt années d'existence, une nouvelle collection, la série "K", le CIRB veut démontrer concrètement le virage qu'il est en train de réaliser au niveau de son orientation générale et surtout le désir de répondre à la sollicitation quotidienne qu'il reçoit de notre société contemporaine plaçée devant de nouveaux besoins linguistiques.

La collection "K" est consacrée exclusivement aux ouvrages produits dans le cadre du projet RELAI (Recherche en linguistique appliquée à l'informatique), projet financé conjointement par l'Université Laval et la compagnie IBM. De plus, cette collection acceptera de publier toute recherche qui poursuit les mêmes objectifs que ceux du projet RELAI et qui s'intègre dans ce qu'il est convenu de dénommer aujourd'hui *les industries de la langue.*

Il est bon de rappeler les objectifs généraux poursuivis par le projet RELAI:

1) Développer un réseau national et international d'information et de documentation dans le secteur de la linguistique appliquée à l'informatique;
2) Analyser et évaluer au point de vue langagier les logiciels de gestion, de technique, de jeux et les didacticiels;
3) Développer et tester des stratégies (code et protocole) d'élaboration de logiciels respectant les contraintes langagières et les comportements culturels des usagers.
4) Analyser dans le contexte de la communication entre l'utilisateur et l'écran, diverses composantes linguistiques du message en vue d'une éventuelle amélioration et de la traduisibilité et du comportement (productivité) de l'utilisateur de l'informatique.

Les nouveaux objectifs du CIRB intègrent ceux du projet RELAI. En effet, le CIRB veut construire un domaine pluridisciplinaire dont l'objectif principal est "d'analyser dans un premier temps, les particularités inhérentes aux situations linguistiques pour un territoire donné, de proposer, dans un second temps, une position politique en matière linguistique enchassée dans un cadre juridique et de déterminer enfin, toutes les applications pratiques dans la vie quotidienne des citoyens que ce soit pour les fonctions de la langue, de l'éducation, dans le monde du travail, de l'administration publique, du commerce, des communications".

Lorne LAFORGE
Directeur du CIRB

Octobre 1987

RÉSUMÉ

Ce logiciel, appelé LEMMATISEUR[c], est un outil d'analyse de textes. Il permet d'analyser les mots dans leur contexte immédiat, afin de constituer une base de données textuelles enrichie par une analyse partiellement automatisée.

LEMMATISEUR permet un travail interactif là où l'intervention humaine est nécessaire. De plus, afin d'économiser sur les interventions répétitives, il comporte un mode d'opération automatique. Celui-ci utilise les données d'analyse accumulées préalablement dans un dictionnaire.

L'ensemble fonctionne de façon conviviale, étant commandé par une hiérarchie de menus couvrant toute la gamme des choix, à partir des classes générales de traitement jusqu'aux opérations détaillées pour les données. Le document présent est consacré, après une introduction qui fait l'apologie de l'approche adoptée, à la description systématique de l'utilisation du logiciel.

[c] Tous droits réservés - Université Laval - 1987.

REMERCIEMENTS

Les travaux qui ont conduit à cette publication ont bénéficié de deux subventions de recherche, en 1985 et en 1986, en provenance du Conseil de recherche en sciences humaines du Canada (CRSHC).

La rédaction du texte a été rendue possible grâce à l'appui du projet RELAI (dans le cadre de l'entente de collaboration entre la Société IBM Canada et l'Université Laval), et l'encouragement de son directeur, Conrad Ouellon.

TABLE DES MATIÈRES

0. Introduction 1

0.1 Un logiciel pour l'analyse des corpus textuels 1
0.2 L'analyse automatique vs. l'analyse interactive 2

1. Présentation de LEMMATISEUR 5

1.1 La diffusion de LEMMATISEUR 5
1.2 Les collaborateurs au développement de LEMMATISEUR 5
1.3 Le contexte matériel 6
1.4 Les logiciels 6
1.5 Le lancement du système 6
1.6 Vue d'ensemble des opérations 7

2. Description des opérations 9

2.1 Traitement préliminaire du corpus 9

OPTION 01: ÉDITEUR 9
OPTION 02: Segmentation du corpus 11
OPTION 03: Transfert d'un corpus sur disque rigide 13

2.2 Génération et maintien de la banque de données 13

OPTION 01: Génération de la banque de données et de ses index 14
OPTION 02: Correction du corpus 15
OPTION 03: Correction du dictionnaire 16
OPTION 04: Mise à jour de la banque des mots exclus 18
OPTION 05: Mise à jour de la banque de données quant aux mots exclus 19
OPTION 06: Mise à jour de la banque de données quant aux corrections apportées au dictionnaire 19
OPTION 07: Mise à jour de la table des catégories grammaticales 22
OPTION 08: Mise à jour des index de la banque de données 22
OPTION 09: Mise à jour des index du dictionnaire 23
OPTION 10: Mise à jour de la banque des mots protégés 23

2.3 Lemmatisation du corpus 24

OPTION 01: Lemmatisation automatique 25
OPTION 02: Lemmatisation interactive 26
OPTION 03: Lemmatisation complémentaire, niveau lexical 30
OPTION 04: Lemmatisation complémentaire, niveau contextuel 31

2.4 Présentation des résultats 35

OPTION 01: Index des mots 36
OPTION 02: Concordance (liste) 36
OPTION 03: Impression du corpus (liste du texte) 38
OPTION 04: Impression du dictionnaire (liste) 39
OPTION 05: Transfert de la banque sur bande magnétique 39
OPTION 06: Transfert de la banque sur disquette(s) 40

0. INTRODUCTION[1]

0.1 Un logiciel pour l'analyse des corpus textuels

Le logiciel LEMMATISEUR a été développé afin de répondre à des besoins croissants d'analyse de données textuelles venant du secteur des sciences humaines; et cela à une époque où les contraintes budgétaires mettaient fin à la possibilité de développer un système particulier pour chaque usage. Il se voulait un outil général pour la constitution de bases de données textuelles.

LEMMATISEUR offre un nombre de fonctions qui facilitent l'analyse formelle d'un corpus donné. Ainsi, il crée une base de données qui peut être enrichie par des catégories d'analyse déterminées selon les besoins de l'utilisateur.

Tous les types d'analyse (linguistique, littéraire, historique, philosophique, juridique, politique, etc...) sont potentiellement admissibles, en autant que l'analyse souhaitée se rapporte à des mots (ou à des unités semblables) dans un corpus de textes.

En guise d'exemple très général, l'utilisateur peut traiter ses données selon les phases suivantes:

1. Saisir les données directement sur clavier à l'aide du logiciel; ou les transmettre par modem ou sur disquette;
2. épurer les données, les corriger et effectuer la segmentation du texte en mots;
3. générer et maintenir les fichiers de la base de données: fichier des occurrences, dictionnaire et leurs index;
4. analyser les mots du corpus à traiter en leur assignant des catégories, des formes lemmatiques ou d'autres éléments d'information;
5. tirer des index, concordances et copies des données à partir de la base.

Le logiciel se veut un outil "transportable" de traitement des corpus textuels. Il a été développé comme logiciel d'application du progiciel dBASE III, un système de gestion de bases de données commercialisé par Ashton Tate Inc. L'utilisateur peut faire appel à toutes les ressources de dBASE III dans l'exploitation des fichiers dans la base qu'il aurait créée.

LEMMATISEUR a été conçu pour fonctionner sur un matériel informatique de grande série. Il peut être implanté sur un PC d'IBM ou ses clones, avec 320 Ko de mémoire vive et deux unités de disquettes de 360 Ko, dans le contexte du système d'opération PC/DOS 2.0. Cependant, pour un rendement satisfaisant, on recommande un PC/AT avec disque rigide et DOS 3.0 ou ses descendants.

Les fonctions centrales sont celles de l'analyse des occurrences textuelles. Les analyses ont pour objectifs l'assignation d'un lemme pour l'identification lexicale de chaque mot-forme, d'où le nom du logiciel, ainsi que l'étiquettage des fonctions, grammaticales ou autres, tel qu'illustré par les exemples du tableau qui suit:

MOT-FORME	LEMME	ÉTIQUETTE 1	ÉTIQUETTE 2
vais	aller	verbe	1 sing. ind.
l	le	article	défini
l	la	article	défini
le	le	pronom	comp. dir.
bureau	bureau #1	nom	"meuble"
bureau	bureau #2	nom	"pièce"
bureau	bureau #3	nom	"agence"

[1]Le texte de l'introduction est une adaptation d'un article paru en 1986: "Le lemmatiseur: un outil interactif pour les données textuelles" dans Langues et linguistique, no. 12, pp 255-268, Université Laval, Québec.

Le résultat recherché est l'obtention d'une base de données enrichie par les catégories d'analyse. Les catégories sont déterminées par l'utilisateur, et peuvent porter sur tous les niveaux d'analyse linguistique: phonétique, morphologique, lexical, syntaxique ou sémantique. Elles peuvent également être d'un ordre autre que linguistique selon la discipline qui intéresse l'analyste: littérature, histoire, sciences politiques, droit, philosophie, etc. En autant que les traits pertinents peuvent être attribués avec une certaine régularité aux éléments constitutifs du texte que sont les mots-formes, le lemmatiseur pourra faciliter le travail en le rendant partiellement automatique. Dans tous les cas (assignation de lemme ou assignation de fonction) nous parlerons de lemmatisation.

LEMMATISATEUR facilite le travail de deux façons: en rendant automatique la partie répétitive du travail, et en travaillant de façon interactive avec l'utilisateur là où il faut une décision humaine. En mode interactif, le logiciel affiche à l'écran, selon certains choix offerts à l'utilisateur, le mot-occurrence à analyser dans son contexte. Il affiche également les choix d'analyse qui ont déjà été acceptés pour le mot-forme et consignés dans un dictionnaire cumulatif, ou il en impose un nouveau, avant de passer au cas suivant. La succession des cas est choisie parmi les options suivantes:

1. une seule occurrence;
2. toutes les occurrences de tel mot-forme;
3. toutes les occurrences dans l'ordre du texte.

En mode de traitement "différé", les occurrences sont analysées automatiquement selon les choix présents dans le dictionnaire. L'utilisateur peut reprendre en mode interactif le résidu d'occurrences non-identifiées. Simultanément, le dictionnaire est augmenté par l'inclusion de chaque nouvelle identification. Ou bien, l'utilisateur peut reprendre les occurrences reconnues comme ambigûes du fait que le mot-forme correspond à deux identifications possibles dans le dictionnaire. Lors du traitement en différé l'occurrence reçoit automatiquement l'identification qui est prioritaire dans le dictionnaire; il reste donc à parcourir les occurrences du mot-forme ambigu afin de relever les ambigüités. Les cas de mots-formes non-identifiés et de mots-formes ambigus peuvent être scrutés séparément avec des options à l'intérieur du mode interactif. Chaque fois qu'un cas fait l'objet d'une décision, il cesse d'être considéré comme ambigu ou non identifié. Avec les options dont il dispose, l'utilisateur peut limiter ses interventions aux seuls mots qui l'intéressent ainsi qu'à ceux qui causent problème.

0.2 L'analyse automatique vs. l'analyse interactive

Si l'on exclut ce qui dépend de l'intervention humaine, LEMMATISATEUR est limité de façon stricte à la distinction des mots-formes. Pourtant, certains systèmes vont beaucoup plus loin dans l'automatisation de l'analyse des textes. Il est possible d'intégrer l'analyse morphologique pour reconnaître les formes déclinées. L'analyse syntaxique automatique permet de faire appel aux contraintes formelles imposées par la langue à ce niveau, et ainsi de mieux cerner le contenu. Pourquoi ne pas incorporer le maximum d'automatisation afin de réduire au minimum la part de travail humain? Certains projets suivent cette voie avec succès.

Pourtant la linguistique automatique n'offre pas la solution totale qui permettrait d'éliminer le recours humain. Il suffit de considérer n'importe quel cas d'ambigüité homographique, comme le mot *porte* qui peut être nom ou verbe. Dans un premier temps, l'analyse automatique du contexte immédiat devrait réduire les ambigüités: *porte* est verbe après un pronom et nom après un article. Il faut cependant regarder plus loin, du fait que *la* et *le* peuvent être pronom ou article. On est donc amené à une analyse syntaxique qui se doit d'être complète afin d'éliminer les cas récalcitrants.

Les travaux d'intelligence artificielle nous montrent jusqu'où il faut pousser les analyses automatiques pour résoudre les ambigüités au niveau des mots-formes. Dans la phrase *Le soldat ferme la porte*, l'ambigüité de *porte* comme verbe (le soldat "solide" porte une chose) ou nom (l'interprétation évidente) résiste à toute analyse syntaxique. Seul le contexte plus large du discours ou des connais-

sances extra-linguistiques permettent de trancher entre les deux sens. Dans une analyse ascendante (en partant des mots vers la phrase), les niveaux inférieurs restent partiellement en plan en attendant l'apport des niveaux supérieurs. Certaines ambigüités se relaient ainsi jusqu'au-delà des limites de la linguistique, là ou l'intelligence artificielle propose de les traiter en faisant intervenir des formalisations des connaissances de l'univers. Autrement dit, à la limite, il faut savoir "de quoi ça parle" avant de savoir "ce que ça dit".

La solution automatique totale est donc une illusion: la compréhension automatique des textes dans sa généralité ne fait pas partie de l'avenir prévisible. Par conséquent, une analyse syntaxique complète et univoque est également impossible. Il reste néanmoins des approches partielles: une analyse des multiples possibilités, même rarissimes, de chaque phrase; une analyse univoque comportant des erreurs; ou une analyse limitée à ce qui est sûr. La première approche implique une révision manuelle complète; la deuxième entraîne un taux d'erreur indéterminé; la dernière offre la possibilité de réduire le nombre d'ambiguïtés qu'il faut réviser.

L'analyse syntaxique automatique comporte un coût qu'il faut peser contre ses bénéfices. Les ressources physiques et humaines requises sont telles que les chercheurs individuels ou en petite équipe ne peuvent songer à développer un système valable. Seuls les centres avec une certaine envergure doivent s'y aventurer. Un corollaire de l'envergure des systèmes est celui de leur inertie. Il faut prendre la grammaire qui est disponible même si elle ne convient pas aux orientations théoriques du chercheur, ou bien il faut élaborer son propre ensemble de règles. Pour un chercheur dont la syntaxe automatique n'est pas la spécialité, cela représente un déplacement important de son attention.

Il y a des facteurs humains qui freinent l'adoption des solutions fortement informatisées. Le chercheur hésite à s'engager dans une voie qu'il ne contrôle pas, ou qu'il craint de ne pas pouvoir contrôler. Pour la personne qui ne travaille pas déjà avec les ordinateurs, il faut ajouter l'appréhension devant cet inconnu, surtout s'il s'agit d'un gros équipement central avec ses couches multiples de logiciels d'exploitation, de personnel et de sa structure bureaucratique. Une solution plus douce, même si elle implique plus de travail humain, devrait être mieux reçue.

LEMMATISEUR se situe clairement parmi les logiciels de traitement *assisté* par l'ordinateur. Il se distingue par sa facilité d'utilisation. À l'instar des tendances à la convivialité des systèmes en micro-informatique, le logiciel fonctionne par menu. L'utilisateur n'a qu'à choisir les fonctions voulues et à fournir ses analyses en réponse aux choix affichés sur l'écran. Il n'a à maîtriser ni le DOS ni le langage dBASE III.

La caractéristique la plus notable pour les utilisateurs en sciences humaines, c'est le contrôle direct sur l'état des données que lui confère l'interactivité du traitement. L'automatisation des processus peut avoir comme conséquence désavantageuse que l'intéressé ne sait pas s'il y a des erreurs dans les résultats, et il est impuissant à en faire la vérification de façon sélective. Le mode automatique de lemmatisation peut très bien engendrer des identifications erronées, mais le mode interactif fournit le moyen de les vérifier et de les corriger. Une décision quant à la catégorisation d'un mot-forme n'empêche pas la reprise, au besoin, de chacune des occurrences du mot-forme en fonction de son contexte particulier.

Le contrôle direct des analyses implique un coût en temps humain. Si l'absence d'erreur est exigée de la part du chercheur, le travail humain devient un investissement indispensable. Comme nous l'avons vu, la lemmatisation entièrement automatique entraîne un taux d'erreurs qui est en général indéterminable.

Si LEMMATISEUR n'élimine pas complètement le travail répétitif, il a quand même l'avantage de le réduire, tout en le faisant porter le plus près possible de l'objet d'intérêt du chercheur. Ce dernier peut limiter son analyse des occurrences à celles qui présentent un intérêt possible. Le nombre de ses interventions est en fonction du nombre d'occurrences qui posent des problèmes d'identifi-

cation. Et chaque intervention porte sur un élément des données du chercheur; non pas sur un code arbitraire dans un langage de contrôle étranger à ses préoccupation principales. Autrement dit, le travail du chercheur entraînera une connaissance accrue de ses données.

Par exemple, il serait possible de réduire le taux d'ambiguïté à l'aide d'une fonction de recherche contextuelle. La distinction entre *le* pronom et *le* article est relativement simple dans un bon nombre de cas. Mais pour des mots-formes moins fréquents que *le*, la rentabilité d'une "désambiguisation" automatique s'estompe. La diminution du travail de vérification des occurrences en contexte serait contrebalancée par le travail d'élaboration des règles de désambiguisation.

Il ne faut pas perdre de vue que le travail d'analyse contextuelle du langage a l'avantage d'être naturel et facile pour les humains, contrairement au travail d'informatisation des procédés d'analyse. Il suffit de penser à la facilité avec laquelle nous réglons des cas comme *Le soldat ferme la porte chaque fois que quelqu'un l'ouvre*. Nous, êtres humains, intégrons à notre lecture du texte une analyse descendante, allant du sens vers les éléments constitutifs. Dans l'état actuel des choses nous sommes beaucoup mieux pourvus que ne le sont les systèmes informatiques pour la compréhension des textes, à tel point qu'il faut bien y penser avant de remplacer les assistants de recherche par un tel système.

Il ne faut pas comprendre par cette plaidoirie qu'il n'y a pas d'intérêt à dépasser le niveau des mots-formes dans le traitement des textes par ordinateurs; au contraire. Tout dépend des objectifs de la recherche. D'autre part, lorsque les objectifs comprennent la reconnaissance d'éléments de contenu moins formels, il peut être stratégiquement plus efficace d'anticiper les problèmes de reconnaissance en faisant la lemmatisation. Par exemple, l'analyse syntaxique automatique est grandement simplifiée si toutes les ambiguités lexicales sont relevées d'avance. En général, lorsqu'il s'agit d'analyse de textes, si les catégories d'analyse peuvent être associées aux mots-formes, une partie seulement du travail peut être automatisée. Il est nécessaire de faire appel aux pouvoirs d'intégration des connaissances que possèdent les humains. Idéalement, un système interactif permet précisément cela, en utilisant à bon escient à la fois les qualités de l'homme et celles de la machine.

1. PRÉSENTATION DE *LEMMATISEUR*

Ce chapitre réunit des sujets qui concernent le logiciel de façon globale, ou qui sont préliminaires à son utilisation.

1.1 La diffusion de *LEMMATISEUR*

Le logiciel LEMMATISEUR est la propriété de l'Université Laval. Il est disponible sur demande auprès de Michael Mepham du Département de langues et linguistique de l'Université Laval. L'Université Laval ainsi que son Centre de traitement de l'information déclinent toute responsabilité lors de l'utilisation du logiciel.

A. L'amélioration progressive des logiciels a pour effet de rendre périmé la documentation imprimée. Afin de pallier à cette lacune, la description des opérations de LEMMATISEUR (le chapitre 2) sera mise à jour lors des modifications de programmation, et la version courante du texte sera disponible sur disquette en accompagnement de celle du logiciel.

1.2 Les collaborateurs au développement de *LEMMATISEUR*

Le logiciel LEMMATISEUR a été élaboré par le Centre de traitement de l'information (CTI) de l'Université Laval. L'équipe de réalisation du CTI se compose comme suit:

Conrad Bourdon	-	chef de projet
Marc Boivin	-	analyste concepteur
Mario Toupin	-	analyste
Denis Légaré	-	programmeur
Lyne Huot	-	programmeure
Nicole Roy	-	programmeure
Gaétane Simoneau	-	programmeure

Le développement du logiciel a été entrepris avec la collaboration du Département de langues et linguistique et plus spécifiquement dans le cadre du projet RELAI (entente IBM-Laval). L'initiative est venue de Michael Mepham, directeur de plusieurs travaux qui ont servi de terrain d'épreuve pour le perfectionnement du système. Plusieurs assistants de recherche ont contribué au processus de rétroaction en tant qu'utilisateurs du logiciel:

Jacques Ladouceur
Danielle Lepage
Richard Bérubé
Mario Plamondon
Ruth Vander Stelt
Beckey Jenkins

Dans ce contexte, le rôle de Jacques Ladouceur mérite une mention spéciale.

Il convient de signaler un projet de la Faculté de théologie qui a également servi de terrain d'épreuve au logiciel. Pierre Chérix et Régine Charron, chercheur et chercheure au projet Nag Hammadi, ont exerçé un rôle analogue à celui des utilisateurs du Département de langues et linguistique.

1.3 Le contexte matériel

LEMMATISEUR fonctionne dans le contexte matériel suivant:

- une unité centrale avec microprocesseur 16 bits, du type PC/XT/AT d'IBM ou compatible;
- un clavier;
- un moniteur;
- une mémoire vive (RAM) de 320 Ko ou plus (LEMMATISEUR en version compilée occupe 220 Ko);
- une imprimante graphique facultative; le code ASCII est étendu à 256 caractères selon le mode IBM;
- une mémoire périphérique minimale de deux unités de disquettes de 360 Ko; de préférence: disquettes de 1.2 Mo ou disques rigides;
- unité de cartouche (bande magnétique) facultative.

1.4 Les logiciels

Tel que diffusé, LEMMATISEUR est logé sur deux disquettes de 360 Ko. L'une des disquettes contient les programmes sous forme compilée, ainsi que tous les fichiers requis pour sa mise en marche. La deuxième disquette contient un fichier de documentation de forme "read me", qui présente toutes les indications utiles du document présent.

Parmi les fichiers de la disquette principale, un certain nombre sont indispensable au fonctionnement de l'une ou l'autre des options du logiciel. Ils sont listés ci-dessous pour fins de vérification. Il faut se garder de les éliminer en faisant le ménage parmi les fichiers qui auraient été crées et emmagasinés sur le même support. Il est prudent de définir une route ("path") pour les systèmes afin de ne pas mélanger les fichiers crées avec ceux qui sont essentiels au système.

Les fichiers indispensables sont:

COD-GRAM.DBF
INDX-STR.DBF
EXCL-STR.DBF
MODI-STR.DBF
SENS-STR.DBF
DICT-STR.DBF
CONPARAM.DBF
IND-STR.DBF
DICT-TEM.DBF

Le programme principal étant compilé, il n'est pas nécessaire de posséder le progiciel dBASE III. Cependant, ce dernier donne beaucoup de possibilités de manipulation des fichiers créés par LEMMATISEUR.

L'utilisateur doit fournir son propre système d'exploitation (DOS). Afin de faciliter l'accès à LEMMATISEUR, il peut constituer une disquette de travail en faisant formater une disquette en disquette système (voir le manuel du DOS), et en faisant un double du logiciel sur cette dernière. Il est prudent de garder la disquette initiale en lieu sûr, et de se servir de la disquette de travail.

1.5 Le lancement du système

L'utilisateur doit fournir une commande de démarrage comportant deux paramètres afin de spécifier la localisation du dictionnaire et du fichier des mots exclus. Cette commande qui peut être intégrée à un fichier à commande directe (fichier avec l'extension .BAT) prend la forme suivante:

[d:]>Lemme [d:] nom du dictionnaire [d:] nom du fichier des exclusions

Ex: A>Lemme B:Dict-lem A:Exclu

N.B. le nom d'un fichier ne peut excéder 30 caractères. Si l'un ou l'autre des deux noms de fichier n'est pas fourni, le système les sollicitera par voie de menu.

En réponse à la commande de démarrage, LEMMATISEUR affiche son logo à l'écran et attend le signal de l'utilisateur avant de présenter le premier d'une série de menus.

1.6 Vue d'ensemble des opérations

LEMMATISEUR est commandé par menus successifs. Il suffit de choisir les options désirées. En guise d'exemple, voici le menu général selon sa présentation à l'écran:

****** LEMMATISEUR **** menu général**
(01) Traitement préliminaire du corpus
(02) Génération et maintien de la banque de données
(03) Lemmatisation du corpus
(04) Présentation des résultats
Entrer le numéro correspondant à votre choix:____
(Quitter: PgDn ou RETOUR)

Chaque option mène à un autre menu de choix plus détaillés conduisant à l'application d'une des fonctions. Nous verrons au chapitre suivant la description de chacune d'entre elles. Pour cette étape-ci, nous faisons un survol de l'ensemble des opérations possibles. Les groupes de choix ci-dessous correspondent aux quatre options du menu général.

Groupe 01: Traitement préliminaire du corpus

1: Editeur de texte pour la saisie du corpus au clavier;
2: Segmentation du corpus en mots-formes;
3: Transfert d'un corpus sur disque rigide.

Groupe 02: Génération et maintien de la banque de données

1: Génération d'une banque de données et de ses index;
2: Correction du corpus;
3: Correction du dictionnaire;
4: Mise à jour du fichier des mots exclus;
5: Mise à jour des mots exclus dans la banque de données;
6: Mise à jour de la banque de données quant aux corrections apportées au dictionnaire (inopérant, à venir);
7: Mise à jour de la table des catégories grammaticales;
8: Mise à jour des index de la banque de données;
9: Mise à jour des index du dictionnaire;
10: Mise à jour de la banque des mots protégés.

Groupe 03: Lemmatisation du corpus

1: Lemmatisation automatique;
2: Lemmatisation interactive;
3: Lemmatisation complémentaire, niveau lexical;
4: Lemmatisation complémentaire, niveau contextuel.

Groupe 04: Présentation des résultats

1: Index des mots;
2: Concordance;
3: Impression du corpus;
4: Impression du dictionnaire;
5: Transfert de la banque sur bande magnétique;
6: Transfert de la banque sur disquette(s).

Enfin, chaque menu indique la façon de sortir du menu courant afin de retourner au menu précédent. Dans le cas présent, pour quitter le menu général du progiciel et retourner au DOS, il faut toucher PgDn ou RETOUR.

2. DESCRIPTION DES OPÉRATIONS

- Nous procédons selon le cheminement des menus.

2.1 Traitement préliminaire du corpus

****** LEMMATISEUR **** menu général**

(01) Traitement préliminaire du corpus
(02) Génération et maintien de la banque de données
(03) Lemmatisation du corpus
(04) Présentation des résultats

Entrer le numéro correspondant à votre choix: 0
(Quitter: PgDn ou RETOUR)

- Pour activer l'option 01, taper 1; un second menu s'affiche comme suit:

******LEMMATISEUR **** Traitement préliminaire écran T1**

(01) Editeur
(02) Segmentation du corpus
(03) Transfert d'un corpus sur disque rigide

Entrer le numéro correspondant à votre choix: _
(Menu précédant: PgDn ou RETOUR)

- Ce menu (écran T1) présente trois options. Toujours de la même façon, entrer le choix.
- Pour sortir de ce menu et retourner au menu général, taper PgDn ou RETOUR.

OPTION 01: ÉDITEUR (écran T1)

Description

a) Ce programme est un éditeur de textes pour la saisie d'un corpus au clavier. Il est surtout utile pour traiter de courts échantillons de texte.
b) Il permet la saisie d'une seule ligne de texte à la fois comprise dans un champ de 70 caractères possibles.
c) Il consacre un champ de huit caractères à la référence de chaque ligne.
d) Ce programme reconnaît trois délimiteurs afin de faciliter la segmentation du texte en mots:

1: le délimiteur d'occurrence est l'espace (code ASCII 032); Ex: *La neige tombe.* (3 occurrences distinctes)
2: le délimiteur de sous-occurrence est l'astérisque (*) (code ASCII 042); il provoque la séparation des composantes accolées d'une expression; Ex: *poly*clinique (polyclinique)* (2 occurrences)
3: le délimiteur de ponctuation est le trait oblique (/) (code ASCII 047); il sert de délimiteur entre le mot et la ponctuation; Ex: *L/'arc/-en/-ciel disparaît/.* (5 occurrences) Sans trait oblique, la ponctuation serait assimilée au mot, pour donner deux occurrences comme ceci: Ex: *dit-il* (une occurrence) Ex: *aujourd'hui* (une occurrence).

e) Le programme procède à la segmentation automatique des occurrences de chacune des lignes de texte.
f) Le programme emmagasine le fichier d'occurrences de mots en format dBase.

Fonctionnement

- Choisir l'option 01 de l'écran T1; l'écran T11 s'affiche comme suit:

****** ÉDITEUR ****** **écran T11**

Nom de la banque _____

DÉFINITION DES DÉLIMITEURS

Délimiteur d'occurrence:
Délimiteur de sous-occurrence:*
Délimiteur de ponctuation: /

- Cet écran demande de préciser le code de l'unité de disque suivi du nom de la banque de données dans un champ de 30 caractères;

Ex: B:corpus (ou C:corpus) (ou Corpus25) (RETOUR)

- Ensuite l'écran T12 s'affiche comme suit:

****** EDITEUR **** saisie** **écran T12**

Référence:_____

Entrez l'information

_____

- Cet écran demande d'entrer un code de référence (8 caractères possibles). Ex: Référence: *00000001* (ou *1*) (RETOUR)

- Ensuite entrer une ligne de texte (70 caractères possibles). Ex: *La lune disparut sous d'épais nuages.* (RETOUR)

- Dans l'espace entre la référence et l'endroit où il faut entrer l'information, l'écran présente toujours la ligne précédente avec sa référence, telles qu'entrées.

- Pour sortir de cette fonction, faire PgDn; ceci ramène l'utilisateur à l'écran T1.

OPTION 02: Segmentation du corpus (écran T1)

Description

a) La fonction de ce programme est de segmenter un corpus de format DOS pour créer un fichier segmenté en format DOS. Elle ne sert pas à traiter les échantillons saisis à l'aide de l'éditeur (option 01).

Voici les caractères reconnus par le programme de segmentation comme séparateurs d'occurrences:

1. l'espace blanc
2. les points: <.> <!> <?> <;> <:>
3. la virgule <,>
4. l'apostrophe <'>
5. les guillemets <">
6. le trait d'union <->
7. les crochets <[> <]>
8. les parenthèses <(> <)>
9. les accolades <{> <}>

b) Les trois points <...> ne sont pas reconnus comme tels; ils sont traités comme trois signes distincts.

c) À cause de leur inclusion dans la liste des séparateurs d'occurrences, l'apostrophe <'> et le trait d'union <-> provoquent la segmentation de chaînes de caractères ordinairement traitées comme des occurrences simples:

Ex: aujourd'hui (deux occurrences)
Ex: c'est-à-dire (quatre occurrences)

d) Les caractères spéciaux qui ne sont pas reconnus comme séparateurs d'occurrences peuvent être utilisés dans un texte pour forcer la concaténation de deux mots:

Ex: aujourd/hui (une occurrence)

e) Lors de la segmentation, les caractères en majuscules sont convertis en minuscules.

f) Lors de la segmentation, le programme peut prélever une chaîne de 1 à 8 caractères qui sert de référence au début ou à la fin (position fixe) de chaque ligne de texte. Cette référence est assignée à chaque occurrence dans la ligne.

g) Le fichier des mots protégés constitue une banque de mots pouvant être sélectionnés à l'avance et segmentés en un seul bloc. Ce fichier doit être créé avant de commencer la segmentation (voir la partie 2.2, option 10).

Fonctionnement

- Choisir l'option 02 de l'écran T1; l'écran suivant se présente comme suit:

```
**** LEMMATISEUR ****segmentation     T21

Quel est le nom du corpus à segmenter?
Quel est le nom du fichier des mots protégés?
Quel est le nom du fichier des occurrences?
Position de début du texte à segmenter?          1
Position de début de la zone référence?          0
Longueur de la zone référence?                   0
```

Description des paramètres pour la segmentation

Paramètre du corpus à segmenter (il doit être de format DOS):
Entrer le nom du fichier à segmenter.

Ex: B:CORPUS (ou CORPUS25) (RETOUR)

Paramètre du fichier des mots protégés:
Entrer le nom du fichier comprenant la banque des mots protégés.

Ex: B:MOTPRO (ou MOTPRO) (RETOUR)

Paramètre du fichier des occurrences:
Entrer le nom du fichier résultant de la segmentation. Ce fichier des occurrences sera de format DOS; ainsi il faudra lui donner l'extension *.TXT* comme suit:

Ex: B:MOTS01.TXT (ou MOTS01.TXT) (RETOUR)

Paramètre de la position de début du texte à segmenter:
Entrer un nombre correspondant à la position du premier caractère du texte à segmenter (par défaut: *1*):

Ex: 1 (ou 10) (RETOUR)

Paramètre de la position de début de la zone référence:
Entrer un nombre correspondant à la position du premier caractère de la zone référence (par défaut: *0*):

Ex: 0 (ou 72) (RETOUR)

Paramètre de la longueur de la zone référence:
Entrer un nombre correspondant à la longueur de la zone occupée par la référence (par défaut: *0*):

Ex: 0 (ou 8) (RETOUR)

- Pendant la segmentation, l'écran ci-dessous s'affiche comme suit:

VEUILLEZ PATIENTER

Segmentation en cours...

- À la fin de la segmentation, le système retourne à l'écran T1.

OPTION 03: Transfert d'un corpus sur disque rigide

Description

Cette fonction permet le transfert d'une banque de données dans le répertoire désiré d'un disque rigide.

Fonctionnement

- Choisir l'option 03 de l'écran T1; l'écran T31 s'affiche comme suit:

****** TRANSFERT ****** **écran T31**

Nom de la banque à transférer sur disque rigide:_____
Nom du répertoire recevant la banque:_____

(Menu précédent: PgDn)

- Entrer le lecteur et le nom de la banque de données. Ex: *B:Corpus* (RETOUR)
- Entrer le nom du répertoire qui recevra la banque de données. Ex: *C:Corpus* (RETOUR)
- Le système retourne à l'écran T1.

2.2 Génération et maintien de la banque de données

****** LEMMATISEUR **** menu général**

(01) Traitement préliminaire du corpus
(02) Génération et maintien de la banque de données
(03) Lemmatisation du corpus
(04) Présentation des résultats

Entrer le numéro correspondant à votre choix: 0
(Quitter: PgDn ou RETOUR)

- Pour activer l'option 02, taper 2; un second menu s'affiche comme suit:

**** Lemmatiseur ** génération et maintien écran G1**

Génération	Correction
(01) Banque de données et ses index	(02) Corpus
	(03) Dictionnaire

Mise à jour
(04) Banque des mots exclus
(05) Banque de données quant aux mots exclus
(06) Banque de données quant aux corrections apportées au dictionnaire
(07) Tables des catégories grammaticales
(08) Index de la banque de données
(09) Index du dictionnaire
(10) Banque des mots protégés

Entrer le nom de votre banque de données: _____
Entrer le numéro correspondant à votre choix: 0
(Menu précédent: PgDn)

- Ce menu (écran G1) présente 10 options. Il faut entrer le code du lecteur suivi du nom de la banque de données. Ex: *B:Corpus* (RETOUR)
- Ensuite, entrer le numéro du choix.
- Pour retourner au menu général, taper PgDn.

OPTION 01: Génération de la banque de données et de ses index (écran G11)

Description

a) Cette fonction génère un nouveau fichier des occurrences avec ses index, en partant d'un corpus à traiter en format DOS. On peut reconnaître les fichiers générés par la forme de leurs noms, comme suit:

fichier DOS du corpus	**xxxxxxxx.TXT**	**Ex: Corpus25.TXT**
fichier des occurrences	**xxxxxxxx.DBF**	**Ex: Corpus25.DBF**
index selon le mot	**xxxx-occ.NTX**	**Ex: Corp-occ.NTX**
index selon la position	**xxxx-no.NTX**	**Ex: Corp-no.NTX**

b) Avant de procéder à l'utilisation d'une nouvelle banque de données, il faut s'assurer de la présence d'un dictionnaire et de ses index. Si l'on veut commencer avec un nouveau dictionnaire, vide au départ, il faut le générer avec l'option 09 (mise à jour des index du dictionnaire).

Fonctionnement

- Choisir l'option 01 dans l'écran G1; l'écran G11 s'affiche comme suit:

***** Lemmatiseur *** Conversion écran G11**

Conversion d'un fichier DOS en fichier DBASE III
et création des index.

Entrer le nom du fichier de format DOS: _____

(Menu précédent: PgDn ou RETOUR)

- Cet écran permet d'entrer le nom du fichier en format DOS pour la conversion. Ex: *B:Corpus25.TXT*
- Ensuite, le système retourne à l'écran G1.

OPTION 02: Correction du corpus (écran G21)

Description

a) Cette fonction permet de modifier le texte avant de procéder à la lemmatisation.

b) Cette fonction efface toute trace de lemmatisation antérieure sur le mot modifié; il faut donc reprendre sa lemmatisation.

c) Il est à conseiller de regénérer les index de la banque de données (option 08) après avoir effectué des corrections, sinon certains éléments corrigés sont traités par le système seulement après les autres; autrement dit, leur traitement se fait comme s'ils étaient ajoutés à la fin du texte. Ceci peut être un inconvénient, mais ne constitue aucunement une erreur de procédure.

Fonctionnement

- Choisir l'option 02 dans l'écran G1; l'écran G21 s'affiche comme suit:

***** Lemmatiseur *** Correction écran G21**

1: lire 2: modifier 3: enlever
4: ajouter 5: continuer 6: sortir

Code d'opération? 0

Mot? _____(30 caractères) Ponctuation:
Numéro de la ligne? _____ Numéro de la référence:

(Exécution de l'opération: PgDn)

- Entrer le code d'opération correspondant au choix, de 1 à 6. (Voir la description ci-dessous.)

- Ensuite entrer le mot à rechercher pour la correction (30 caractères maximum), si ce mot n'est pas précisé, un message signalera cet oubli.

- Ensuite entrer s'il y a lieu le signe de ponctuation qui suit le mot. Si ce signe n'est pas précisé, il sera ignoré dans la recherche.

- Ensuite entrer le numéro de ligne où se situe le mot (8 caractères), si ce numéro n'est pas précisé, le premier cas dans l'ordre du texte sera traité.

- Enfin, entrer le numéro de référence. Si ce numéro n'est pas précisé, le premier cas, dans l'ordre des références du texte, sera traité.

- Pour activer l'opération de correction, taper PgDn.

- Pour retourner à l'écran G1, choisir l'option 6.

DESCRIPTION DES CODES D'OPÉRATIONS:

1: lire: Cette opération doit précéder toutes les autres. Elle trouve l'occurrence demandée selon son numéro de ligne et l'affiche en contexte.
2: modifier: Cette opération permet de remplacer l'occurrence affichée en entrant un autre mot et la ponctuation qui l'accompagne dans l'espace prévu à cette fin.
3: enlever: Cette opération permet d'enlever l'occurrence affichée. Cependant, dans le cas où cette occurrence est suivie d'une ponctuation, il faudra la reporter sur le mot précédent.
4: ajouter: Cette opération ajoute un mot à la suite de l'occurrence affichée en contexte. Il faudra reporter la ponctuation qui l'accompagne à la suite de ce nouvel ajout.
5: continuer: Cette option donne la possibilité de répéter l'opération précédente. Ex: Si l'opération précédente était de LIRE le mot SOLEIL, cette opération irait lire le prochain mot SOLEIL dans le corpus. Le principe est le même pour les autres codes d'opérations.
6: sortir: Cette option permet de suspendre la fonction et de revenir à l'écran G1.

OPTION 03: Correction du dictionnaire (écran G32)

Description

Cette fonction permet:

a) d'inscrire un lemme pour un mot absent du dictionnaire;
b) d'ajouter un lemme pour un mot présent au dictionnaire;
c) d'enlever un lemme pour un mot;
d) de changer l'ordre dans la liste des lemmes pour un même mot. Ceci permet de fixer le choix prioritaire (le premier) lorsqu'un mot a plus d'un lemme dans le dictionnaire.

Fonctionnement

- Choisir l'option 03 dans l'écran G1; l'écran G32 s'affiche comme suit:

```
*** Lemmatiseur ***        Correction        écran G32

Quel est le mot à mettre à jour?_____________________

(Menu précédent: PgDn ou RETOUR)
```

- Entrer à l'écran le mot à mettre à jour dans le dictionnaire. Ex: Quel est le mot à mettre à jour? *soleil* (RETOUR)
- Si le mot n'est pas présent au dictionnaire, l'écran G33 s'affiche, comme ci-dessous, avec le message suivant en haut de l'écran (avec le mot "soleil" comme exemple):

```
L'occurrence (soleil) est absente au dictionnaire.
```

- Pour inscrire un lemme au dictionnaire, taper le lemme vis-à-vis le code de sélection 11, suivi de RETOUR.
- Taper sur la même ligne le code grammatical, suivi de RETOUR, puis taper 11 à la fin de la ligne suivante afin d'imposer le choix numéro 11.
- Si le mot est présent au dictionnaire, l'écran G33 s'affiche comme suit:

```
*** Lemmatiseur ***                              écran G33

soleil
^
code lemme(s)          code gramm.      code grammatical
  1: soleil                 10          10:  nom
                                        20:  pronom
                                        30:  conjonction
                                        40:  verbe
                                        50:  adverbe
                                        60:  préposition
                                        70:  coordonnant
                                        80:  interjection
                                        90:  autre

11: soleil                  0

** Entrer le code correspondant au lemme désiré  0
   Le code 99 permet de suspendre la session.
     (Exécution de l'opération: PgDn)
```

- Pour ajouter un nouveau lemme au(x) lemme(s) déjà existant(s), ajouter un nouveau code grammatical à la ligne 11 et inscrire 11 à la ligne suivante. Cette fonction peut modifier l'orthographe d'un lemme.

Ex: le mot SOLEIL est inscrit dans le dictionnaire avec le code grammatical 10 et on veut lui ajouter un lemme supplémentaire (code gramm. 90). Il faut procéder comme suit:

1: l'écran G32 demande le mot à mettre à jour: taper SOLEIL suivi de RETOUR.
2: l'écran suivant (G33) présente le mot recherché, son lemme et son code grammatical (code de sélection 1 de l'écran G33).
3: à la ligne 11, le mot-lemme est affiché, il faut insérer à cette ligne le nouveau code grammatical et taper 11 à la ligne suivante. Le nouveau lemme est ajouté au dictionnaire.

- Pour enlever un lemme, ou en changer l'ordre de priorité, faire deux RETOUR successifs vis-à-vis le code de sélection 11, puis inscrire à la ligne suivante le code de sélection du lemme à traiter. Ensuite, l'écran G31 s'affiche comme suit:

***** Lemmatiseur *** Correction écran G31**

(01) Enlever le lemme
(02) Modifier sa priorité

Entrer le code correspondant à votre choix: 1
(Menu précédent: PgDn ou RETOUR)

- Pour enlever le lemme: taper 1.
- Pour changer l'ordre de priorité d'un lemme: taper 2.
- L'écran G33 reviendra s'afficher avec les modifications complétées.
- Pour sortir de l'écran G33 et retourner à l'écran G1, taper 99.

OPTION 04: Mise à jour de la banque des mots exclus (écran G41)

Description

a) Cette fonction permet de spécifier les mots qu'on ne désire pas lemmatiser. Ces mots sont inscrits dans le fichier des mots exclus et servent à marquer les occurrences identiques dans le fichier des occurrences (voir l'option 5 ci-dessous).
b) La liste d'exclusion est utile surtout pour écarter des mots fréquents qui n'auraient pas de pertinence pour l'étude; les articles, par exemple.

Fonctionnement

- Choisir l'option 04 dans l'écran G1; l'écran G41 s'affiche comme suit:

*** Lemmatiseur *** mode d'exclusion écran G41

Quel est le mot à exclure? __(30 caractères)__

(Menu précédent: PgDn ou RETOUR)

- Entrer le mot à exclure (30 caractères ou moins). (RETOUR)
- L'écran G41 redemande après chaque mot exclu le prochain mot à exclure. Pour terminer la session et retourner à l'écran G1, ne rien entrer et faire un RETOUR.

OPTION 05: Mise à jour de la banque de données quant aux mots exclus (écran G51)

Description

a) Cette fonction marque dans le fichier des occurrences celles qui sont à exclure de la lemmatisation.
b) Il faut faire cette mise à jour lorsque de nouveaux mots ont été ajoutés au fichier des mots exclus (l'option 04).

Fonctionnement

- Choisir l'option 05 dans l'écran G1; l'écran G51 s'affiche comme suit:

*** Lemmatiseur *** génération et maintien écran G51

VEUILLEZ PATIENTER

Mise à jour des mots exclus de la banque
de données en cours d'exécution

- Après le traitement, le système retourne à l'écran G1.

OPTION 06: Mise à jour de la banque de données quant aux corrections apportées au dictionnaire

Description

a) Chaque correction faite au dictionnaire (option 03) est inscrite dans un fichier des corrections. Ce fichier comporte ainsi une liste des modifications qui ont été faites aux entrées du dictionnaire, avec la date de l'opération.
b) Cette option permet d'une part de gérer le fichier des corrections, en retirant de la liste celles dont on s'est déjà occupé.
c) D'autre part, cette option permet de faire répercuter dans les banques de données déjà lemmatisées, les modifications qui auront été faites dans le dictionnaire depuis la lemmatisation des banques.

Fonctionnement

- Choisir l'option 06 dans l'écran G1;
  l'écran G61 s'affiche comme suit:

***** Lemmatiseur *** génération et maintien écran G61**

****** Dictionnaire actif: b:dict-lem ******

(01) Traiter la liste des retraits
(02) Apporter les modifications au(x) corpus

Entrer le numéro correspondant à votre choix: 0
(Menu précédent: PgDn ou RETOUR)

- Ce menu (écran G61) présente trois options. Toujours de la même façon, entrer le choix.
- Pour sortir de ce menu et retourner à l'écran G1, PgDn ou RETOUR.

CHOIX 01: Traiter la liste des retraits

Description

a) Cette fonction regroupe trois traitements concernant les corrections apportées au dictionnaire, ce sont:

option 01: Impression de la liste de retraits.
Cette fonction permet l'impression d'une liste de retraits dans le fichier des corrections apportées au dictionnaire, avec en-tête comprenant la date de l'impression, le titre, le nom du dictionnaire concerné. La liste comprend quatre colonnes dont, successivement, une pour le mot, une pour le numéro de lemme, une pour le code grammatical et la dernière pour la date de modification où le retrait a été effectué.

option 02: Suppression des modifications du XXXX-XX-XX au XXXX-XX-XX.
Cette fonction permet le retrait dans le fichier des entrées pour les corrections, des modifications apportées entre deux dates précises. Dans l'écran G621 (voir Fonctionnement, après l'option suivante) un message s'affiche dès que cette option est activée demandant la date de début et celle de la fin de la période à supprimer. Pour annuler l'opération, laisser les champs blancs.

option 03: Suppression du fichier de modifications de ce dictionnaire.
Cette fonction permet la destruction des modification du dictionnaire. Dans l'écran G621 (voir ci-après) un message va s'afficher comme suit:

Confirmer la destruction du fichier de corrections (O/N):

Fonctionnement

- Taper 1 dans l'écran G61;
  l'écran G621 s'affiche comme suit:

***** Lemmatiseur *** génération et maintien écran G621**

(01) Impression de la liste de retraits
(02) Suppression des modifications du XXXX-XX-XX au XXXX-XX-XX
(03) Suppression du fichier des modifications de ce dictionnaire

Entrer le numéro correspondant à votre choix: ___
(Menu précédent: PgDn ou RETOUR)

- Taper le numéro de votre choix.
- Pour retourner à l'écran G61, PgDn ou RETOUR.

CHOIX 02: Apporter les modifications au(x) corpus

Description

a) Cette fonction permet la mise à jour des banques de données quant aux modifications contenues dans le fichier des corrections.
b) Dans une séance, l'utilisateur peut spécifier un maximum de 10 fichiers des banques de données à corriger. S'il y en a plus, il faut répéter l'opération.
c) Il faut bien noter dans chaque champ la bonne structure des répertoires et sous-répertoires.

Fonctionnement

- Taper 2 dans l'écran G61;
  l'écran G631 s'affiche comme suit:

***** Lemmatiseur *** génération et maintien écran G631**

Format des noms de corpus **D:\SOUS-DIR.EXT\BANQUEXX.EXT**

(28 caractères

Entrer un caractère pour débuter la mise à jour: ___
(Menu précédent PgDn ou RETOUR)

- Chaque champ peut contenir vingt-huit caractères.
- Pour débuter la fonction, taper une touche quelconque.
- Pour revenir à l'écran G61, PgDn ou RETOUR.

OPTION 07: Mise à jour de la table des catégories grammaticales (écran G71)

Description

a) Cette fonction sert à changer les dénominations des catégories grammaticales dans la table de l'écran G33 (on la retrouve partiellement encadrée à droite de l'écran).
b) Chacune des catégories peut être nommée selon les besoins.
c) Chaque nom de catégorie occupe un champ de 20 caractères, il faut donc une expression de 20 caractères ou moins.

Fonctionnement

- Choisir l'option 07 dans l'écran G1; l'écran G71 s'affiche comme suit:

*** Lemmatiseur *** génération et maintien écran G71

DÉFINITION DES CATÉGORIES GRAMMATICALES

10: nom	60: préposition
20: pronom	70: coordonnant
30: conjonction	80: interjection
40: verbe	90: autre
50: adverbe	

(Menu précédent ou exécution de l'opération: PgDn)

- Taper dans les espaces appropriés les nouvelles dénominations suivies de RETOUR.
- Les dénominations sont automatiquement modifiées dans la table de l'écran G33.
- Après le traitement, le système revient à l'écran G1.

OPTION 08: Mise à jour des index de la banque de données (écran G81)

Description

a) Cette fonction regénère les index du fichier des occurrences.
b) La regénération des index de la banque est à conseiller lorsque des opérations de correction (option 02) ont été effectuées sur le corpus, ceci afin d'assurer que les fichiers d'index retrouvent une organisation simple et efficace.

Fonctionnement

- Choisir l'option 08 dans l'écran G1; l'écran G81 s'affiche comme suit:

***** Lemmatiseur *** génération et maintien écran G81**

VEUILLEZ PATIENTER

Mise à jour des index de la banque de données en cours d'exécution

- Après le traitement, le système revient à l'écran G1.

OPTION 09: Mise à jour des index du dictionnaire (écran G91)

Description

a) Cette fonction regénère les seuls index du dictionnaire. Cette opération est recommandée lorsque le nombre de corrections qui ont été apportées au dictionnaire risque de complexifier les index au point de ralentir leur fonctionnement.
b) S'il n'y a pas de dictionnaire dans le répertoire approprié, cette fonction en créera un avec le nom spécifié lors du lancement du programme *(voir Vue d'ensemble des opérations)*.

Fonctionnement

- Choisir l'option 09 dans l'écran G1; l'écran G91 s'affiche comme suit:

***** Lemmatiseur *** génération et maintien écran G91**

VEUILLEZ PATIENTER

Mise à jour des index de la banque de données en cours d'exécution

- Après le traitement, le système revient à l'écran G1.

OPTION 10: Mise à jour de la banque des mots protégés (écran G101)

Description

a) Cette fonction crée le fichier des mots protégés s'il n'existe pas, ou, s'il existe déjà, permet la modification de son contenu.
b) Ce fichier sert exclusivement au processus de segmentation (voir *2.1 Traitement préliminaire du corpus*, option 02). Si ce fichier n'est pas créé ici, la segmentation ne pourra fonctionner.

Fonctionnement

- Choisir l'option 10 dans l'écran G1; l'écran G101 s'affiche comme suit:

***** Lemmatiseur ***** Mots protégés écran G101

Quel est le nom du fichier des mots protégés? _____

- Entrer le code de l'unité de disque suivi du nom du fichier choisi;

Ex: B:MOTPRO (ou C:MOTPRO) (ou MOTPRO) (RETOUR)

- Ensuite l'écran G102 s'affiche comme suit:

***** Lemmatiseur ***** Mots protégés écran G102

1- AFFICHER
2- AJOUTER DES MOTS
3- CORRIGER UN MOT
4- TRIER
5- FIN

CHOIX ? __

- Entrer le chiffre correspondant à l'option désirée (RETOUR);
- Pour sortir de cette fonction, taper 5, et le système revient à l'écran G1.

2.3 Lemmatisation du corpus

**** LEMMATISEUR **** menu général

(01) Traitement préliminaire du corpus
(02) Génération et maintien de la banque de données
(03) Lemmatisation du corpus
(04) Présentation des résultats

Entrer le numéro correspondant à votre choix: 0
(Quitter: PgDn ou RETOUR)

- Pour activer l'option 03, taper 3; un second menu s'affiche comme suit:

***** Lemmatiseur *** lemmatisation écran L1**

(01) Lemmatisation automatique
(02) Lemmatisation interactive
(03) Lemmatisation complémentaire, niveau lexical
(04) Lemmatisation complémentaire, niveau contextuel

Entrer le nom de votre banque de données:_____
Entrer le numéro correspondant à votre choix: ____
(Menu précédent: PgDn)

- Ce menu demande, en premier lieu, le nom de la banque de données suivi de RETOUR; Ex: *B:Corpus* (RETOUR)
- À la ligne suivante faire un choix parmi les quatre modes de lemmatisation.
- Pour revenir au menu général: PgDn.

OPTION 01: Lemmatisation automatique (écran L11)

Description

a) Cette fonction traite, en mode différé, chaque entrée dans le fichier des occurrences (à l'exception de celles qui sont énumérées dans le fichier des mots exclus) en l'identifiant au mot-forme comparable dans le dictionnaire.
b) Pour chaque occurrence dont la forme se trouve dans le dictionnaire, une relation d'identification à cette forme est établie. Ceci équivaut à l'assignation d'une catégorie, d'un lemme, d'une précision lexicale si elle existe dans le dictionnaire (voir l'option 03 de l'écran L1). La catégorie du mot-forme est également assignée directement à l'occurrence dans le fichier des occurrences.
c) Les occurrences identifiées à l'aide du dictionnaire sont marquées comme lemmatisées; celles dont la forme n'est pas présente dans le dictionnaire conservent leur statut de mot non-lemmatisé.
d) S'il y a plus d'un lemme pour une même occurrence, le lemme prioritaire (le premier dans l'ordre) lui sera assigné. La situation d'ambiguïté sera signalée pour traitement ultérieur (option 02, ci-dessous).

Fonctionnement

- Dans l'écran L1, entrer le nom de la banque de données suivi de RETOUR; Ex: *B:Corpus* (RETOUR).
- Ensuite choisir l'option 01 (taper 1); l'écran L11 s'affiche comme suit:

VEUILLEZ PATIENTER

Lemmatisation de la banque de données en cours d'exécution.

- Un message, s'il y a lieu, avertit de l'absence de mots au dictionnaire. L'utilisateur doit donc les inscrire par lemmatisation interactive (voir l'option suivante).
- Après l'opération, on retourne à l'écran L1.

OPTION 02: Lemmatisation interactive (écran L21)

Description

a) Cette option traite en mode interactif les occurrences afin que l'utilisateur puisse intervenir dans l'assignation des données d'analyse.
b) La lemmatisation interactive se fait selon quatre sous-options (voir ci-dessous). Dans chaque cas, l'occurrence est affichée en contexte (écran G33).
c) Les données d'analyse fournies par l'utilisateur sont ajoutées ou bien au dictionnaire, ou bien au fichier des occurrences.

Fonctionnement

- Dans l'écran L1, entrer le nom de la banque de données suivi de RETOUR.
- Ensuite choisir l'option 02 (taper 2); l'écran L21 s'affiche comme suit:

***** Lemmatiseur *** lemmatisation écran L21**

(01) Lemmatisation de tous les mots n'ayant pas de lemme
(02) Lemmatisation de tous les mots ambigus
(03) Lemmatisation d'un mot
(04) Lemmatisation de toutes les occurrences d'un mot
(05) Suspension de la lemmatisation

Entrer le code correspondant à votre choix: 0

- Ce menu présente cinq choix dont les quatre premiers sont les possibilités de lemmatisation interactive.
- Entrer le code de votre choix, selon le cas.
- Pour suspendre la lemmatisation et retourner à l'écran L1, entrer 05.

CHOIX 01: Lemmatisation de tous les mots n'ayant pas de lemme (écran G33)

Description

a) Cette fonction permet de reprendre la lemmatisation de toutes les occurrences dont la forme est absente du dictionnaire.
b) Les cas sont présentés à l'écran dans l'ordre de leur apparition dans le corpus d'origine.

Fonctionnement

- Taper 1 dans l'écran L21; l'écran G33 s'affiche comme suit (les caractères soulignés constituent un exemple):

```
*** Lemmatiseur ***                          écran G33

      ligne:   1                  référence: 00000001
la lune s'est levée juste avant le soleil
                 ^

code lemme(s)          code gramm.      code grammatical
                                        10: nom
                                        20: pronom
                                        30: conjonction
                                        40: verbe
                                        50: adverbe
                                        60: préposition
                                        70: coordonnant
                                        80: interjection
                                        90: autre
11: juste                   50

** Entrer le code correspondant au lemme désiré  11
   Le code 99 permet de suspendre la session.

      (Exécution de l'opération: PgDn)
```

- L'occurrence identifiée par le signe "^" est présentée en contexte avec son numéro de ligne et sa référence.

- Les catégories principales sont affichées sous la rubrique "code grammatical", avec leur dénomination par défaut (voir l'option 07 de l'écran G1 pour modifier l'appellation des catégories).

- Vis-à-vis le code de sélection 11, l'utilisateur inscrit son choix de lemme (maximum 20 caractères avec RETOUR) et de code grammatical. Ce dernier peut prendre n'importe laquelle valeur de 0 à 99.

- À la ligne suivante, il faut inscrire 11.

- Afin d'activer l'assignation des données dans les fichiers d'occurrences et de dictionnaire, faire PgDn.

- Pour suspendre la séance et retourner à l'écran L21, taper 99 à l'endroit désigné.

CHOIX 02: Lemmatisation de tous les mots ambigus (écran G33)

Description

a) Cette fonction permet de lemmatiser toutes les occurrences de mots ambigus tels que marqués lors d'une lemmatisation automatique.
b) Les cas sont présentés à l'écran dans l'ordre de leur apparition dans le corpus d'origine.
c) Les occurrences traitées perdent leur statut de cas ambigus.

Fonctionnement

- Taper 2 dans l'écran L21; l'écran G33, s'affiche comme suit (les caractères soulignés constituent un exemple):

```
*** Lemmatiseur ***                          écran G33

     ligne:        1          référence: 00000001
la lune s'est levée juste avant le soleil
*

code lemme(s)        code gramm.      code grammatical
  1: *la               90           10:nom
  2: la                20           20:pronom
                                    30:conjonction
                                    40:verbe
                                    50:adverbe
                                    60:préposition
                                    70:coordonnant
                                    80:interjection
                                    90:déterminant
 11: la                20

**   Entrer le code correspondant au lemme désiré 01
     Le code 99 permet de suspendre la session.
          (Exécution de l'opération: PgDn)
```

- L'occurrence est présentée comme au choix précédent (choix 01), avec, en plus, les lemmes et les codes grammaticaux possibles (codes de sélection 1 et 2)

- Le lemme qui a déjà été assigné, par défaut ou par choix, est signalé par un astérisque (code de sélection 1).

- L'utilisateur peut inscrire un autre choix au code de sélection si les choix existants ne conviennent pas (maximum de 10 choix par mot-forme).
- À la ligne suivante, l'utilisateur indique son choix (01, 02 dans l'exemple, ou 11 s'il veut ajouter une autre possibilité). Le choix par défaut est inscrit d'avance à la ligne 11.
- Faire PgDn et la marque d'ambiguïté est annulée.
- Pour suspendre la séance et retourner à l'écran L21, taper 99 sous la ligne 11.

CHOIX 03: Lemmatisation d'un mot (écrans L22 et G33)

Description

a) Cette fonction permet de lemmatiser une seule occurrence d'un mot, ce qui est utile surtout pour corriger une erreur.
b) L'occurrence présentée est la première dans l'ordre alphabétique en commançant à la ligne spécifiée par l'utilisateur. Si aucune ligne n'est spécifiée, la première occurrence du mot dans le corpus est présentée.

Fonctionnement

- Taper 3 dans l'écran L21; l'écran L22 s'affiche comme suit:

***** Lemmatiseur *** lemmatisation écran L22**

Quel est le mot à lemmatiser? _____

A quelle ligne le mot apparaît-il? ____,____0
(Menu précédent ou exécution de l'opération: PgDn)

- Cet écran demande le mot à lemmatiser. Ecrire le mot suivi de RETOUR.
- On peut inscrire le numéro de ligne pour préciser la recherche; ensuite, faire un RETOUR. En l'absence du numéro, la recherche se fera à partir du début.
- L'écran G33, comme pour les choix précédents, s'affiche et fonctionne de façon semblable.
- Pour suspendre la session et revenir à l'écran L21, taper 99 dans le bas de l'écran.

CHOIX 04: Lemmatisation de toutes les occurrences d'un mot (écrans L22 et G33)

Description

a) Cette fonction permet de lemmatiser successivement toutes les occurrences d'un mot.

b) Les cas sont présentés à l'écran dans l'ordre de leur apparition dans le corpus, à partir de la ligne spécifiée par l'utilisateur.

Fonctionnement

- Taper 4 dans l'écran L21; ce choix conduit à l'écran L22 (comme pour le choix précédent).
- On doit spécifier le mot et on peut indiquer la ligne.
- Le système passe de l'écran L22 à l'écran G33. Le fonctionnement de ce choix est semblable au choix 01 et 02.
- Pour suspendre la session et revenir à l'écran L21, taper 99 sous la ligne 11.

CHOIX 05: Suspension de la lemmatisation

Fonctionnement

- Taper 5 dans l'écran L21 pour retourner à l'écran L1.

OPTION 03: Lemmatisation complémentaire, niveau lexical (écran L31)

Description

a) Cette fonction permet d'ajouter des données d'analyse à chaque occurrence d'un mot du corpus. Les occurrences sont traitées successivement dans l'ordre d'apparition dans le corpus.
b) Les données occupent un champ de 30 caractères sans structure inhérente. Il incombe à l'utilisateur de définir un format et une codification appropriés à ses besoins.
c) L'utilisateur peut choisir un champ de précision pour une occurrence parmi les possibilités déjà reconnues pour le lemme, ou en ajouter une nouvelle. Il y a une limite de 20 précisions différentes pour chaque lemme dans le dictionnaire.

Fonctionnement

- Dans l'écran L1, entrer le nom de la banque de données suivi de RETOUR.
- Choisir l'option 03; l'écran L31 s'affiche comme suit:

***** Lemmatiseur *** précision écran L31**

Quel est le mot à préciser? _____

(Menu précédent: PgDn ou RETOUR)

- Taper le mot à préciser.
- Ensuite l'écran L32 s'affiche comme suit (les caractères soulignés constituent un exemple):

```
*** Lemmatiseur ***        précision        écran L32

     Ligne:   1            référence: 00000001
la lune s'est levée juste avant le soleil

No. Précision                    No. Précision
1. sens: se réveiller            2. sens: se mettre debout


21: _____(30 caractères)_____

**  Entrer le no. correspondant au sens désiré  0
    Le code 99 permet de suspendre la session.

          (Exécution de l'opération: PgDn)
```

- Dans l'exemple (caractères soulignés), il y a 2 précisions préexistantes.
- Entrer une nouvelle précision à la ligne 21 au besoin. Chaque précision doit avoir 30 caractères ou moins. (RETOUR)
- Taper 21 à la ligne sous-jacente pour une nouvelle précision, ou le numéro approprié pour une possibilité existante (exemple: *1* ou *2* dans l'écran L32).
- Ensuite, la fonction continue avec l'occurrence suivante. Si la précision reste la même, taper deux RETOUR de suite.
- Pour suspendre la session, taper 99 vis-à-vis du numéro. On retourne à l'écran L31.
- Pour retourner à l'écran L1, taper PgDn ou RETOUR.

OPTION 04: Lemmatisation complémentaire, niveau contextuel (écran L41)

Description

a) Cette fonction permet d'assigner des codes d'analyse à chaque occurrence et de les conserver dans le fichier des occurrences.
b) La fonction prévoit un seul champ non structuré de 20 caractères par occurrence. La codification du champ est laissée entièrement au gré de l'utilisateur.

Fonctionnement

- Dans l'écran L1, entrer le nom de la banque de données suivi de RETOUR.

- Choisir l'option 04; l'écran L41 s'affiche comme suit:

```
*** Lemmatiseur ***        précision        écran L41

(01) Précision de toutes les occurrences non précisées
(02) Précision d'un mot
(03) Précision de toutes les occurrences d'un mot
(04) Précision de toutes les occurrences rattachées à une
     catégorie grammaticale
(05) Suspension de la lemmatisation

Entrer le code correspondant à votre choix:  0
```

- Ce menu présente cinq choix dont les quatre premiers sont des modalités de lemmatisation complémentaire.
- Entrer le code de votre choix, selon le cas.
- Voici la description et le fonctionnement de chacune des options:

CHOIX 01: Précision de toutes les occurrences non précisées (écran L43)

Description

a) Cette fonction présente successivement dans l'ordre du texte toutes les occurrences du corpus qui n'ont pas encore été précisées à ce niveau.
b) Toutes les occurrences sont éligibles, même celles qui n'ont pas été analysées au niveau principal (catégorie et lemme). De cette façon, l'utilisateur peut escamoter le niveau principal et se servir de celui-ci s'il convient mieux à ses besoins.

Fonctionnement

- Taper 1 dans l'écran L41; l'écran L43 s'affiche comme suit (les caractères soulignés constituent un exemple):

```
*** Lemmatiseur ***        précision        écran L43

Ligne:  1          référence: 00000001

la lune s'est levée juste avant le soleil

Quelle est la précision:  (20 caractères)

** Le code 99 permet de suspendre la session.
```

- La précision ne peut excéder 20 caractères.
- Si on n'inscrit pas de précision et qu'on tape un RETOUR, l'opération se poursuit sur la prochaine occurrence non précisée.
- Pour terminer la session, taper 99 suivi de RETOUR, on retourne à l'écran L41.

CHOIX 02: Précision d'un mot (écrans L42 et L43)

Description

a) Cette fonction présente une seule occurrence pour précision, la première en partant de la ligne spécifiée par l'utilisateur.
b) Si l'occurrence a déjà été traitée à ce niveau, les nouveaux codes d'analyse remplacent les premiers.

Fonctionnement

- Taper 2 dans l'écran L41; l'écran L42 s'affiche comme suit:

```
*** Lemmatiseur ***        précision        écran L42

Quel est le mot à préciser? _____________________

A quelle ligne le mot apparaît-il:    0

(Menu précédent ou exécution de l'opération: PgDn)
```

- Taper le mot à préciser suivi de RETOUR.
- Taper le numéro de ligne pour préciser la recherche; sinon cette dernière s'effectuera à travers tout le corpus.
- Ensuite l'écran L43 s'affiche et s'opère comme au choix précédent (choix 01).
- Pour terminer la session, taper 99 suivi de RETOUR.
- Pour retourner à l'écran L41, taper PgDn ou RETOUR.

CHOIX 03: Précision de toutes les occurrences d'un mot (écrans L42 et L43)

Description

a) Cette fonction permet d'analyser successivement dans l'ordre textuel toutes les occurrences d'une même forme.

b) Les nouveaux codes remplaceront les premiers, s'il y en a qui sont déjà assignés.

Fonctionnement

- Taper 3 dans l'écran L41; l'écran L42 s'affiche comme au choix précédent (choix 02).
- Taper le mot à préciser suivi de RETOUR.
- Taper un second RETOUR au lieu de spécifier la ligne (à moins de ne pas vouloir commencer au début du corpus).
- Ensuite l'écran L43 s'affiche et s'opère comme au choix 01.
- Pour retourner de l'écran L42 à l'écran L41: PgDn ou RETOUR.

CHOIX 04: Précision de toutes les occurrences rattachées à une catégorie grammaticale (écran L41 et L43)

Description

a) Cette fonction permet d'analyser successivement, dans l'ordre textuel, toutes les occurrences des formes ayant la même catégorie au niveau principal.
b) À l'aide de cette fonction, l'utilisateur peut effectuer des analyses contextuelles détaillées pour une classe de mots voulue.

Fonctionnement

- Taper 4 dans l'écran L41; un message s'affiche au bas de l'écran L41, comme suit:

Quel est le code de catégorie grammaticale? __(RETOUR)

- La fonction recherche toutes les occurrences qui répondent à cette catégorie grammaticale et les affiche une à la fois dans l'écran L43, qui s'opère comme au choix 02.
- Pour terminer la session, taper 99 suivi de RETOUR.

CHOIX 05: Suspension du traitement

a) Pour suspendre la lemmatisation complémentaire et retourner à l'écran L1: taper 5.

b) Pour retourner au menu général, taper PgDn dans l'écran L1.

2.4 Présentation des résultats

```
****LEMMATISEUR **** menu général

(01) Traitement préliminaire du corpus
(02) Génération et maintien de la banque de données
(03) Lemmatisation du corpus
(04) Présentation des résultats

Entrer le numéro correspondant à votre choix: 0
     (Quitter: PgDn ou RETOUR)
```

- Pour sortir du progiciel et retourner au DOS: PgDn ou RETOUR
- Pour activer l'option (04), taper 4; un second menu s'affiche comme suit:

```
*** Lemmatiseur ***        résultats        écran R1

(01) Index des mots
(02) Concordance
(03) Impression du corpus
(04) Impression du dictionnaire
(05) Transfert de la banque sur bande magnétique
(06) Transfert de la banque sur disquette(s)

Entrer le nom de votre banque de données:_____
Entrer le numéro correspondant à votre choix:  0
     Menu précédent: PgDn)
```

- Ce menu comporte six utilitaires permettant soit l'impression de diverses données (options 01, 02, 03, 04), soit le transfert de la banque de données sur un autre support (options 05 et 06).
- En ce qui a trait aux options 01, 02, 03, 04, dans l'écran R1, le choix de l'un d'entre eux fait suite à un menu qui s'affiche comme suit:

```
*** Lemmatiseur *** Sélection des périphériques-SELECSOR

Si des résultats doivent être produits,
sous quel format désires-vous les obtenir ?

(01) Sur papier
(02) Dans un fichier sur disque
(03) Sur papier et sur disque
(00) Retour au menu précédent

Entrer le numéro correspondant à votre choix: 0

     (Terminer la saisie: PgDn)
```

- Les trois premiers choix permettent la présentation des résultats sous trois formes:

  option 01: impression des données sur papier;
  option 02: transfert des données sur disque;
  option 03: les deux options précédentes sont combinées.

- Pour retourner à l'écran R1, taper 0.

- Voici la description de ces utilitaires:

OPTION 01: Index des mots

Description

a) Cette fonction fait l'impression d'un index de toutes les occurrences de tous les mots d'un corpus.
b) L'index est en ordre alphabétique; d'abord selon les lemmes, et deuxièmement, selon les mots-formes du même lemme, ensuite, en ordre textuel des occurrences pour chaque mot-forme.
c) L'impression des pages se fait avec en-tête comprenant: le titre, la date, le nom de la banque de données et la numérotation de la page.

Fonctionnement

- Activer l'imprimante.

- Dans l'écran R1, entrer le nom de la banque de données suivi de RETOUR.

- Ensuite choisir l'option 01.

- L'écran SELECSOR va s'afficher. Choisir un des trois choix. Dès que sont sélectionnés les choix 02 et 03, deux messages s'affichent au bas du menu, comme suit:

  Entrer le nom du fichier désiré: __ (30 caractères) (RETOUR)
  Désirez-vous conserver les en-têtes sur le texte du fichier?
  (O/N) O

- La liste d'index s'affiche à l'écran en même temps qu'elle s'imprime sur papier.

- Après l'opération, on retourne à l'écran R1.

OPTION 02: Concordance (liste) (écran R21)

Description

a) Cette fonction fait l'impression d'un index d'une classe de mots assortis de leur contexte pour chaque occurrence.
b) La concordance est en ordre alphabétique des mots-formes et en ordre textuel des occurrences pour un même mot-forme.

c) La sélection des mots dans la concordance se fait selon les paramètres d'analyse spécifiés par l'utilisateur.
d) L'impression des pages se fait avec en-tête comprenant: le titre, la date, le nom de la banque de données et la numérotation de la page.
e) La concordance s'affiche à l'écran pendant l'impression.

Fonctionnement

- Activer l'imprimante.
- Dans l'écran R1, entrer le nom de la banque de données suivi de RETOUR. Choisir l'option 02.
- Ensuite l'écran SELECSOR va s'afficher. Comme pour l'option 01, faire un choix et répondre aux messages affichés.
- Un messages supplémentaire va s'afficher comme suit:

Voules-vous effectuer la concordance générale? (O/N) O

Si la réponse est positive, le traitement se fera simultanément. Si elle est négative, l'écran R21 s'affiche comme suit:

*** Concordance *** mot à sélectionner écran R21

Quel est le mot lemme? _____ (RETOUR)
Quel est le mot forme? _____ (RETOUR)
Quelle est la référence? _____ (RETOUR)
Quelle est la catégorie grammaticale? ___ (RETOUR)
Quelle est la précision au niveau lexical? 0 (RETOUR)
Quelle est la précision au niveau contextuel? _____
RETOUR)
Désirez-vous les mots exclus (O/N)? ___ (RETOUR)
Le mot à sélectionner est-il conditionnel au(x) mot(s)
précédent(s) (nombre de mots: 0, 1, 2 ou 3)? 0 (RETOUR)

Le mot à sélectionner est-il conditionnel au(x) mot(s,
suivant(s) (nombre de mots: 0, 1, 2 ou 3)? 0 (RETOUR)

(Menu précédent ou exécution de l'opération: PgDn)

- Après l'opération, on retourne à l'écran R21.
- Pour retourner au menu précédent (écran R1): PgDn.
- Ce menu présente divers choix dont la description détaillée suit:

Description des paramètres de sélection pour la concordance

Paramètre: le mot lemme
En spécifiant un lemme, l'utilisateur obtient toutes les occurrences de toutes les formes d'un même lemme (ex: pour toutes les formes d'un même verbe).

Paramètre: le mot forme
En spécifiant un mot-forme, l'utilisateur obtient les occurrences d'un seul mot forme.

Paramètre: la référence
En spécifiant la référence, seules les occurrences appartenant à la subdivision correspondant au corpus sont retenues.

Paramètre: la catégorie grammaticale
En spécifiant une catégorie grammaticale, seules les occurrences de cette catégorie sont retenues.

Paramètre: la précision au niveau lexical
En spécifiant la précision au niveau lexical, seules les occurrences possédant cette valeur sont retenues.

Paramètre: la précision au niveau contextuel
En spécifiant la précision au niveau contextuel, seules les occurrences possédant cette valeur sont retenues.

Paramètre: les mots exclus
L'utilisateur peut faire inclure ("O") ou exclure ("N") les occurrences des mots qui appartiennent à la liste des mots exclus. Ce choix doit s'ajouter à l'un des précédents.

Paramètre: ...mots précédents ou suivants
L'utilisateur peut imposer une contrainte de cooccurrence contextuelle sur la sélection, en spécifiant le ou les mots formes qui doivent se trouver dans le voisinage. On peut définir un contexte comprenant de 1 à 3 mots devant et de 1 à 3 mots après les occurrences en cause.

OPTION 03: Impression du corpus (liste du texte)

Description

a) Cette fonction fait l'impression du texte avec référence à la ligne.
b) L'impression des pages se fait avec en-tête comprenant: le titre, la date, le nom de la banque de données, la numérotation de la page.
c) On peut imprimer une partie du texte en spécifiant le début et la fin de la section désirée.

Fonctionnement

- Activer l'imprimante.
- Dans l'écran R1, entrer le nom de la banque de données suivi de RETOUR.
- Choisir l'option 03; l'écran R31 s'affiche comme suit:

***** Impression du texte ***** **écran R31**

Quel est l'intervalle des références à imprimer?

de _____(RETOUR) à_____ (RETOUR)

(Exécution de l'opération: PgDn)

- Inscrire l'intervalle des références à imprimer si une partie seulement du texte est visée; sinon faire deux RETOUR et tout le texte est imprimé.
- Le texte s'affiche à l'écran en même temps qu'il s'imprime.
- Après l'opération, on retourne à l'écran R1.

OPTION 04: Impression du dictionnaire (liste)

Description

a) Cette fonction fait l'impression du dictionnaire et rend compte du type, des lemmes et des codes grammaticaux de chaque mot-forme. Le code du type indique si l'entrée dans la liste correspond à un lemme pour lequel il n'y a pas de forme textuelle reconnue (type 1), à un mot-forme de type unique (type 2), ou à un mot-forme à plusieurs lemmes possibles (type 3).
b) L'impression des pages se fait avec en-tête comprenant: le titre, la date, le nom de la banque de données et la numérotation de la page.

Fonctionnement

- Activer l'imprimante.
- Dans l'écran R1, entrer le nom du dictionnaire suivi de RETOUR.
- Ensuite choisir l'option 04.
- La présentation de la liste du dictionnaire s'affiche à l'écran pendant l'impression.
- Après l'opération, on retourne à l'écran R1.

OPTION 05: Transfert de la banque sur bande magnétique

Cette fonction n'était pas encore disponible à la date de publication.

OPTION 06: Transfert de la banque sur disquette(s)

Description

a) Cette fonction permet de copier sur disquette tout le contenu d'une banque de données.
b) La banque peut occuper plus d'une disquette.
c) Cette fonction effectue le transfert uniquement sur l'unité A.

Fonctionnement

- Insérer dans l'unité A la disquette réceptrice.

- Dans l'écran R1, entrer le nom de la banque de données. Faire un RETOUR et la fonction va s'activer. Si le contenu d'une banque de données est supérieur à la capacité d'une disquette, un message indiquera, le moment venu, d'introduire une autre disquette pour compléter le transfert.